Working paper ZMB-24096 July 2025

# Flooding and child health outcomes in Zambia

Jeremy Foltz
Theresa Ng'andu

DIRECTED BY
FUNDED BY

# Flooding and Child Health Outcomes in Zambia *

Theresa Ng'andu
Agricultural and Applied Economics Department,
University of Wisconsin-Madison

July 28th, 2025

## 1 Abstract

I study the impact of flooding on the incidence of diseases such as malaria, cholera, typhoid, and dysentery from 2012 to 2020 in children under the age of five years old in Zambia. I use health data from the Zambian Ministry of Health, Health Management Information System (HMIS), and satellite flood data from the NASA MODIS Water Product to estimate a linear probability model with fixed effects at the ward level, the fourth and smallest administrative unit in Zambia. I find that flooded wards have a significantly higher incidence of malaria and cholera in children than non-flooded wards. There is no difference, however, in the incidence of typhoid and dysentery in children between flooded and non-flooded wards. The results are not driven by wards located in main towns or provincial capitals where economic activity and development are high, which shows that the risk of the increased incidence of diseases due to

*I would like to thank the International Growth Centre (IGC) Zambia for funding this project. I am specifically thankful to Thandiwe Ng'ombe and Rory Hardie at the IGC for their help in getting the Ministry of Health data and in shaping the policy aspects of the paper. I also thank the provincial and district health officers and public health officers in various provinces for meeting with me and granting me permission to get data from the provinces and/or districts. I am grateful to Charles Chikwanda from Kakoso Level 1 Hospital in Chililabombwe for facilitating meetings with some district health officers in Ndola, Kitwe, Chingola, and Chililabombwe to discuss how health data is recorded. I am also grateful to my advisor, Dr. Jeremy Foltz, for the academic, financial, and moral support, the Department of Agricultural and Applied Economics, and the University of Wisconsin-Madison for their consistent financial support. I also thank my colleagues in my department, and discussants at the Columbia University IPWSD workshop, and the AERE at WEAI conference for their comments on this project. The findings and conclusions in this paper are my own as the author and do not necessarily represent the views of the University of Wisconsin - Madison and affiliates, the International Growth Centre, or anyone involved in the data collection process.

flooding among children is shared across all flooded wards in Zambia.

## 2 Introduction

Climate change predictions for sub-Saharan African countries indicate heavy precipitation and pluvial flooding. This means that if African countries do not engage in effective city planning as urbanization increases, these extreme weather conditions could be economically costly and hinder sustainable development. As urbanization in the form of infrastructure development and residential housing rises in African cities, the consequences of such rapid growth, such as how increase in impervious surfaces, can increase surface water run-off and contribute to flooding, should be considered.

In Zambia, the urban population growth rate stands at 3.5% [1], which has demanded more infrastructure building in terms of housing, hospitals, and commercial buildings. Some of this residential housing has extended to historical forest reserves and historical conservation areas such as Forest 26 and Forest 27 reserves, and Marshlands by the issuance of presidential instruments to degazette these areas for commercial and residential development. [2] At the same time, Zambia has seen an increase in the occurrence of (flash) floods, which can contribute to the spread of diseases such as malaria and waterborne diseases such as cholera, dysentery, and typhoid. Cholera outbreaks also tend to happen during rainy seasons[3], and health facilities report increases in cases of malaria and waterborne diseases. In this paper, I use a linear probability model with fixed effects to answer the following research question: does flooding increase the incidence of diseases such as malaria, and waterborne diseases such as cholera, typhoid, and dysentery in children under the age of 5 years old in Zambia?

While many studies have shown relationships between rainfall patterns and the prevalence of diseases in African cities such as Dar-es-Salaam in Tanzania (Picarelli et al., 2017), or of flooding and malaria transmission (Boyce et al., 2016), few studies focus on the effects of flooding for an extended period on the prevalence of diseases. Existing literature looks at flooding from a disaster perspective, often focusing on a big flooding event in a particular area or region, and then examining the effects of that event on outcomes such as child immunization rates (Chandir et al., 2023). My research combines aspects of existing research by examining the impact of flooding on child health outcomes, and adds to this literature by looking at flooding over 9 years from 2012 to 2020, and at outcomes for all wards in districts in Zambia. I

[1] According to the 2022 census report, available here: https://www.zamstats.gov.zm/2022-census/reports/

[2] In 2017, 2018, and 2019, former President Edgar C. Lungu de-gazetted parts of a historical forest reserve, Forest 27, to be open for commercial and residential infrastructure by signing Statutory Instruments. The most controversial of the three instruments was that in 2019, which reduced the boundaries of the reserved area from 1764 hectares to less than half, around 716 ha. Forest 27, initially declared a protected area by the former British colony in 1942, was meant to be a water recharge zone, thereby preventing floods when it rains.

[3] The most recent outbreak happened during the 2023/2024 rainy season.

exploit variation in whether a ward, which is the smallest administrative boundary in Zambia, experiences a flood in a month or not, and how many flood events it experiences in a year, to estimate the effects of flooding on the incidence of malaria, cholera, typhoid, and dysentery as measured by the number of cases reported in health facilities in that ward.

In addition to examining flooding and diseases across wards for the entire country, I also use health facility-based data from the Zambian Ministry of Health. This data records disease cases of all individuals who reported to the facility for an illness such as malaria, cholera, typhoid, and dysentery. The advantage of using this data, which comes directly from the Ministry of Health, Health Management Information System (MoH HMIS) database, over other health survey data is that it is more comprehensive and is not based on data recorded over a certain period, but data recorded every month as long as people are visiting the facilities. In analyzing health effects across all wards in Zambia, I offer a rare opportunity to delve deeper into the year-to-year effects of flooding on the incidence of diseases across all areas in Zambia based on whether there is flooding or not in a ward, and on the number of flood events in each of those wards. The policy implication for my research is that policymakers can take measures to mitigate the occurrence of malaria, cholera, typhoid, and dysentery among the most vulnerable during the flooding seasons.

I hypothesize that flooding can increase the spread of malaria and waterborne diseases by providing breeding grounds for pathogens and providing a mechanism for the faster spreading of the diseases through water contamination. Because flooding is non-random and the prevalence of disease is influenced by many factors, I follow both a selection on observables and a selection on unobservables approach. I control for socioeconomic and geographic variables that impact flooding and diseases and use ward and time effects to isolate the effect of flooding on the prevalence of diseases. The ward fixed effects control for ward-specific time-invariant factors such as attitudes and norms that could affect the number of cases of diseases in that ward, and the year fixed effects to control for time-varying factors that affect all wards equally. In another specification, I use constituency by year fixed effects and ward fixed effects to control for year to year changes and local development at the constituency level, as a constituency is the third administrative unit in Zambia, and funds such as constituency development funds (CDF) which are responsible for local community development, are allocated at this level (Musamba & Phiri, 2019). This fixed effects estimation method has been used in similar studies looking at certain activities such as forest loss on diseases, as seen by the paper by (Berazneva & Byker, 2017; Garg, 2019) on the impacts of forest loss on the incidence of malaria and diarrhea diseases in Nigeria, and also on infant mortality (Berazneva & Byker, 2024).

The main result of this paper is that when there is flooding, the probability of there being a cholera case recorded in a facility in a ward increases by about 0.0008 or 0.08 percentage points (whether there is a cholera outbreak or not), representing a percent change of 74 percent when the baseline sample average probability of a ward having a cholera is 0.1 percentage points. For malaria, when it floods, the probability of there being a malaria case recorded in a facility in a given ward is 0.04 or 4 percentage points, representing a 5 percent increase in the probability from the baseline sample average probability of 0.795 or 80 percentage points. These results are consistent with the existing literature regarding floods (or even rainfall) and cholera cases (Boyce et al., 2016; Deshpande et al., 2020; Picarelli et al., 2017), and flooding and disease. Given the total number of malaria confirmed cases reported among children under the age of 5 years between 2012 and 2020 is 14300000[4], a back of the envelope calculation indicates without floods, 57200 of these malaria cases could have been prevented.

For typhoid and dysentery, I do not observe any changes in the probability of a ward recording more cases when it floods versus when it does not flood. There are many reasons why I do not find an effect, but one possible explanation lies in the way that typhoid and dysentery data are recorded in the HMIS database. Some district health officials noted that typhoid and dysentery cases are rarely confirmed with tests, so they are "suspected" or "clinical" cases that are determined based on the symptoms observed and reported at the facility. Thus, some cases of these diseases may also be lumped in with "diarrhea" reported cases in the HMIS database, which I do not include in this analysis.

I also examine if the incidence of diseases due to flooding is driven by main towns in Zambia, and find that this is not the case, i.e., the incidence of diseases is not different between main towns and non-main towns, which shows that the impact of flooding is distributed across all flooded wards, whether the wards are the center of development and economic activity, or not. Overall, the results in this paper offer new insights into the health impacts of flooding.

The rest of this paper is organized as follows: in section III, I go over the existing literature and my contributions to the research on flooding and diseases; in section IV, I describe the data with some summary statistics in the form of graphs; in section V, I outline the empirical specification, challenges to the identification strategy, and the methodology I use in my estimation; in section VI, I present the results and a discussion of the results; and in section VII, I conclude this paper with steps for more analysis.

# 3 Literature Review

The relationship between flooding and diseases is well documented in the public health literature, such as the impact of severe flooding on malaria, and malaria-related hospital visits in Kasese District in Uganda (Boyce et al., 2016), flooding and the breakout of infectious diseases in Kenya (Okaka & Odhiambo, 2018), flash floods and incidence of malaria in Almanagil in

[4]Please note that these are cases, so if a child had malaria three times a year, this was indicated as three cases in the data.

Sudan (Elsanousi et al., 2018). For Zambia, little is known about how flooding impacts the prevalence of diseases in children. Previous studies in public health, such as that by Sasaki et al. (2009), find that rain and poor drainage systems are associated with an increased prevalence of cholera outbreaks in the Lusaka district, and Suhr and Steinert (2022) documents that flood exposure increases in malaria and cholera cases. One thing to note is that these existing studies are localized to one region of a country. In my research, I study flooding for all wards in Zambia, and examine the impact of these floods on the incidence of not just malaria, but also named waterborne diseases such as cholera, typhoid, and dysentery.

In economics, some studies have examined the impact of rainfall shocks on cholera outbreaks in Dar-es-Salaam in Tanzania, using the vulnerability of a place to flooding as a proxy for floods (Picarelli et al., 2017), while some studies have focused on large flood shocks and their impacts on health aspects such as child immunization rates in Pakistan (Chandir et al., 2023). These papers all examine flooding and, in most cases, include some econometric methods to estimate effects, but one common theme is that flooding is treated as a shock rather than something that could occur almost every year. Some research in public health has looked at floods and dysentery in China over long periods (Ni et al., 2014), but studies that look at malaria, and the most common waterborne diseases such as cholera, typhoid, and dysentery, in sub-Saharan Africa are scarce. My research distinguishes itself from existing literature by examining the impact of flooding that is not restricted to disaster-type flooding on the incidence of diseases in children. Thus, I quantify the impact of seasonal flooding coming from yearly (extreme) rainfall rather than flooding from only one big event or shock. My research fills both the period gap and the gap in the variety of diseases examined by going beyond one flooding event, and looking at several diseases beyond malaria and cholera.

My study is also novel in that it uses unique data records with diseases disaggregated by name and age from the Ministry of Health. Most existing studies on this topic of health impacts rely on Demographic and Health Survey (DHS) data (Berazneva & Byker, 2017, 2024; Berggreen & Mattisson, 2024), and where health data from health facilities is available, it is usually present for only a small part of the region (Boyce et al., 2016) over a short period. While the DHS dataset is a household level dataset with a good set of household-level covariates, it is not a yearly survey - data can take up to five years to be collected and made available, which makes it unsuitable for the analysis I conduct in this paper as I aim to look at the year to year impact of floods over 9 years from 2012 to 2020 on child health outcomes. Furthermore, the DHS records cases of "diarrhea", which lumps most waterborne diseases in this category. The data I use distinguishes between the different types of waterborne diseases: from cholera, to typhoid, and dysentery, which then allows me to examine the effect of flooding on the incidence of each of these diseases. Moreover, the malaria data used in this analysis are confirmed cases of malaria, which means that a test was conducted on the individual who visited the health facility, and the malaria case was confirmed, so it is not "fever" being mistaken for malaria.

My work also contributes to the existing research on some of the drivers of malaria in countries in sub-Saharan Africa. There is a vast amount of research on how certain environmental factors, such as forest loss, can lead to a higher incidence of malaria (Berazneva & Byker, 2017; Garg, 2019), and in the case of exposure to forest loss, can also increase infant mortality rates within the first months of birth of the child (Berazneva & Byker, 2024). Other research has also looked at how certain activities, such as gold mining, can increase the number of malaria cases in places where gold mines open (Pagel, 2025). Some studies have focused on heavy rainfall as a driver of diarrhea diseases (Deshpande et al., 2020), while some studies have identified that being close to waterbodies (which is also highly correlated with riverine flooding) increases a district's vulnerability to being a cholera hotspot (Mwaba et al., 2020), and some point to the lack of drainage systems as increasing cholera outbreaks (Sasaki et al., 2009). Heavy rainfall, temperature, and flooding seem to be the most common aspects of the climate that have been associated with the prevalence of waterborne diseases (Levy et al., 2016), and in my study, I focus on flooding, although rainfall is a control in my regressions. The results of my research add to this discussion on flooding and diseases by showing that the impact of flooding extends beyond cholera and malaria to typhoid, a rarely studied waterborne disease in this type of literature.

# 4 Data

## 4.1 Flood Data

The flood data comes from the NASA legacy MODIS Water Product (MWP), which was used to capture flooding at 250$m$ spatial resolution for the years before 2020. The data being used in this analysis is from 2012 to 2020 and covers the entire country of Zambia. These are two-day flood products, which means that if a pixel of water is detected as a flood, then it must have been there the previous day as well. This is an advantage in getting rid of false positives (in addition to the surface water and cloud shadow masking already conducted in the detection of the flood). The MODIS Water Product has the advantage of capturing floods at a smaller spatial resolution, which enables me to also calculate the percentage of an area that is flooded.

The disadvantage of using this MWP flood product may be that it does not capture flash floods, especially the ones bound to happen in cities, and is subject to errors such as underreporting of small floods (i.e., the ones not captured by the media), or floods in remote areas, as documented by the paper by (Patel, 2023). To this extent, the flood data used in this analysis may be on the lower bound for areas with higher levels of impervious surfaces, or for areas that are more likely to experience flash floods. I also use local and online sources on flooding, such as ReliefWeb and Floodlist, to restrict the flooding months to December, January, February, and March in the analyses, as these are the rainy season months and therefore are likely to be

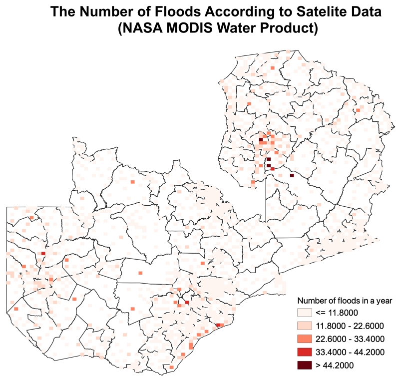


Figure 1: Number of floods

months when the country has flooding (as also reported by ReliefWeb and Floodlist).

A flood, in the MWP flood product, is defined as: "water above a certain level as compared to a reference water map showing normally expected water, such as rivers or lakes." [5]. In the MWP flood product, this flood definition corresponds to a pixel value of 3. A ward is counted as having had a flood if it has pixels of value 3. I measure flood in ward $w$ in month $m$ in year $t$, denoted by $Flood_{wmt}$, as first, a binary variable that equals 1 whenever the ward has pixels of value 3, and zero otherwise. I then measure the total number of flood events in a ward that month and year as the total number of times we have pixel values equal to 3 in that month for that year as the flood data is measured at the daily level, which gives the Flood variable as $Flood_{wt}$ i.e., number of flood events in a ward for a given year. Figure 1 is a map of Zambia with constituency boundaries[6] showing how the number of floods is distributed across the country:

## 4.2 Health Data

The health data come from the Zambian Ministry of Health (MoH) Health Management Information System (HMIS) database, which was obtained after getting ethical clearances in the form of a minimal risk Institutional Review Board (IRB) clearance from the University of Wisconsin - Madison, a local registration and study clearance with the National Health Research Authority (NHRA) in Zambia, and a local ethical clearance from the University of Zambia Biomedical Research Ethics Committee (UNZABREC), and approval letters from each of the provincial health offices in Zambia to be able to get approval from the Ministry of Health Headquarters to get access to the database. From this database (HMIS), I have data on confirmed

[5] MODIS Aqua+Terra Global Flood Product MCDWD$_{L}3_{N}RT distributed from NASA LANCE. Available on-line[https://www.earthdata.nasa.gov/global-flood-product]. DOI: 10.5067/MODIS/MCDWD_{L}3_{N}RT.061$

[6] For map visibility purposes, the map is constructed at the constituency rather than the ward level.

malaria cases [7], cholera cases, dysentery cases, and typhoid cases for children under the age of 5 years old. The data is recorded at the monthly and facility level for every operational health facility in Zambia from 2012 to 2020. This health data is also used to calculate the number of facilities in a ward for any year.

## 4.3 Rainfall, and Temperature Data

The impact of flooding on the prevalence of diseases, especially when certain diseases such as cholera tend to be seasonal, cannot be done without including rainfall and temperature data. For rainfall data, I use the daily Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) as aggregated by the University of California, Santa Barbara Funk et al. (2014) to calculate the mean, minimum, and maximum rainfall for each ward every day, and then take the total rainfall for the month for that year. For temperature, I use the historical mean monthly maximum temperatures, as aggregated by Fick and Hijmans (2017). As with the rainfall data, I follow a similar approach when analyzing the temperature data by taking the mean, maximum, and minimum of the maximum land surface temperature at the ward level for each month and year, and use the maximum land surface temperature as the temperature control for the analysis.

## 4.4 Other Covariates

As this is an analysis on diseases, controlling for population is important to account for places that may have a larger share of children, and places with higher population densities - all factors that can influence the prevalence of diseases in an area. Thus, for population data, for years between 2012 and 2020, I use satellite population estimates from World Pop [8]. These are counts of individuals at the $100 metres$ resolution, estimated from "unconstrained top-down" methods, i.e., uninhabited areas are not masked but counted. As the under-5-year-olds only make up a portion of the population, I take shares at the provincial level from the latest 2022 Census of this age group, and adjust the population estimates to reflect the group. I then calculate the share of the under 5 years old population in each ward as the number of under 5 years old children in that ward for that year, divided by the total population of that ward for that year. I calculate population density as the total number of people in a ward in a given year

[7] Confirmed cases are tested either using blood tests, or rapid diagnostic tests, and differ from clinical cases, which are when the child exhibits symptoms of the disease but is not tested to confirm if the malaria parasite is present in the blood or not.

[8] WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).

divided by the area in square kilometers for that ward.[9].

I include nightlights data as a measure of wealth, or economic activity control for the possibility that some areas might be more well-developed, which could impact the prevalence of diseases, and also be related to the occurrence of floods. These dynamic nighttime lights data are obtained from Li et al. (2021), which are taken at a spatial resolution of 30-arc seconds or $1km$. The harmonized nighttime lights data are a combination of the "Temporally calibrated DMSP-OLS NTL time series data from 1992-2013" and the "Converted NTL time series from the VIIRS data (2014-2021)". Nightlights are a good proxy for economic activity in developing countries, as shown by Chen and Nordhaus (2010). Finally, I include other geographical and topographic factors that can influence the incidence of diseases and the occurrence of flooding. These include the digital elevation data from the Global 30 Arc-Second Elevation (GTOPO30) (1996), and data on proximity to waterbodies, such as the distance to rivers and lakes from the Global Self-consistent, Hierarchical, High-resolution Geography Database.

# 5 Identification and Empirical Strategy

## 5.1 Identification

I use a fixed effects estimation method to isolate the causal impact of flooding on the incidence of diseases. I use the variation in flooding in a ward from year to year to get this estimate. Because other factors affect the the incidence of diseases in any ward over time, I also control for many other variables such as rainfall, temperature, economic activity, distance to waterbodies such as rivers and lakes, log of population count for children under five years in a ward, and elevation of a ward (with regard to the sea level). The main identification challenge is that there could be many more factors beyond the variables controlled for that could be driving the number of cases of diseases observed. This, however, should not be a concern as the variables I control have been used in various similar papers that examine the impact of either rainfall, or flooding on the prevalence of diseases (Berggreen & Mattisson, 2024; Boyce et al., 2016; Deshpande et al., 2020; Picarelli et al., 2017). Thus, the identifying assumption here is that conditional on the variables controlled for, the ward within variation has a causal interpretation of the effect of flooding on the incidence of malaria, cholera, typhoid, and dysentery. Fixed effects have been used in similar studies involving forest loss, such as that by (Berazneva & Byker, 2017, 2024; Garg, 2019).

The second concern stems from the way flooding has been measured, and the fact that satellite flood data has been known to under-count floods in certain parts of the country that may be remote, as evidenced by Patel (2023). From our data, the places most likely affected

[9]The area is taken from the shapefile from the Humdata website: https://data.humdata.org/dataset/zmb-ndvi-subnational

by this are urban areas where flash floods are common, and since the MODIS Water Product only records pixels as floods if they are present for more than two days, this would bias the results downwards for urban areas. To address this issue, I only count floods during the months that are also reported in the news. The other concern, which is related to the cases of disease recorded, is that rural or remote areas may not be as good at recording cases as urban areas due to factors such as having fewer resources (such as working internet and electronic devices) to enter cases and report them to the Ministry of Health Headquarters. While this concern is valid, I do not have any evidence that suggests that rural areas are under-reporting cases, and if this is a factor of concern, it should be controlled for in the measure(s) of development variable(s).

## 5.2 Empirical Strategy

The dependent variable is a binary indicator of whether facilities in a ward recorded cases of malaria, cholera, dysentery, and typhoid, or not. So $Pr(disease = 1)_{wt}$ indicates whether facilities in ward $w$ record a disease case in a given month and year $t$ or not; $Flood_{wt}$ is a dummy variable that equals 1 if there is a flood in a ward in a given month and given year; $Rain_{wt}$ as the total amount of rainfall in $mm$ in a month in a year; the interaction term $Flood \cdot Rain_{wmt}$ which captures how the combination of rain and a flood can impact the incidence of a diseases; $\omega_w$ as the distance to the river; $\gamma_w$ as the distance to the lake; $\theta_w$ as ward fixed effects; $\tau_{ty}$ as year fixed effects; and $\mathbf{X}_{wt}$ as the vector of ward level covariates such as the temperature, elevation, log of population for children under 5 years old, number of facilities in a ward, distance to rivers, distance to lakes, and nightlights. I estimate the following linear probability model, specifically using the:

$$Pr(disease_{wt} = 1) = \beta_0 + \beta_1 \cdot Flood_{wt} + \beta_2 \cdot Rain_{wt} + \beta_3 \cdot Flood_{wt} \cdot Rain_{wt} + \beta_4 \cdot \mathbf{X}_{wt} + \theta_w + \tau_t + \varepsilon_{wt}$$

I present results for both when ward and year fixed effects are used, and when ward and constituency by year fixed effects are used as a robustness check, and because both groups of fixed effects capture different aspects of unobservables.

In some specifications, I also include $\mu_{ct}$ as the constituency-by-year fixed effects (constituency denoted by $c$) in the regression model, and the results are represented in the last column of each results table. In another specification, the variable $Flood_{wt}$ is the number of floods in a given ward in a year. The results of this estimation, where the independent variable $Flood_{wt}$ is a continuous variable, are presented in Tables 6 to 9. In each of the regressions, I cluster the standard errors at the ward level to allow for correlation of errors and allow for autocorrelation within the wards over time.

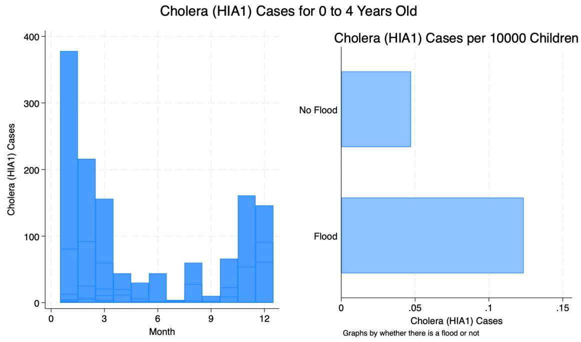


Figure 2: Cholera (HIA1) cases

# 6 Results and Discussion

## 6.1 Summary Statistics

the left-hand side of Figures 2 to 5, we see how the number of cases of each type of disease for children under 5 years old differs by month, with month 1 corresponding to January, and month 12 to December. From these figures, we see that cholera cases are high during the rainy season months of November to March (months 11 to 3), while for the other three diseases, this seasonality pattern is not very clear from the bar graphs. On the right-hand side of each of these figures, we have horizontal bar graphs showing how the number of cases of each type of disease per 10000 of the population of children under 5 years old. Again, there are visible differences in the number of cases between flooded and non-flooded wards, with cholera and malaria cases being higher in flooded than non-flooded wards, while for dysentery and typhoid cases, we see the opposite, with non-flooded areas having a higher number of cases per 10000 of the under 5 years old population than the flooded wards.

In Table 1, I show the descriptive statistics for the covariates, dividing the mean values between flooded and non-flooded wards in any given year. Wards that were flooded at some point in our data have more cases of malaria, cholera, and typhoid per 10000 of the 0 to 4 years old age group population compared to non-flooded wards. This number of cases is, however, reversed for dysentery cases per 10000 of the child population. We also see that flooded areas have higher mean rainfall, lower temperatures, tend to be found at a lower elevation, and have a shorter distance to rivers. We also see that they have less economic activity, are located further from lakes, and have fewer facilities compared to non-flooded wards. Seeing that the mean values between flooded and non-flooded wards are statistically significant, I control for all of these covariates in the regressions.

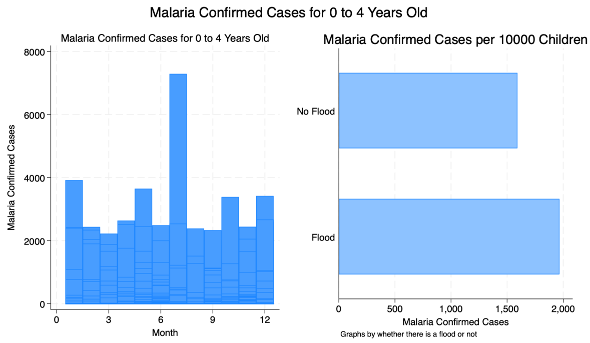


Figure 3: Malaria confirmed cases

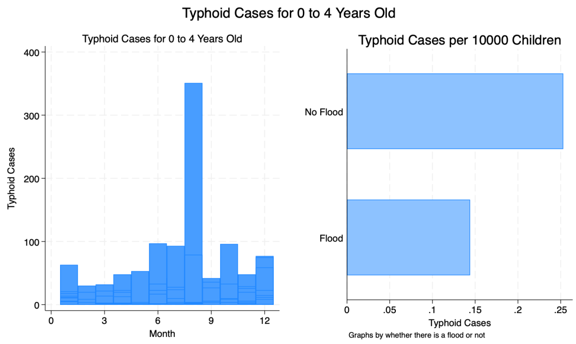


Figure 4: Typhoid cases

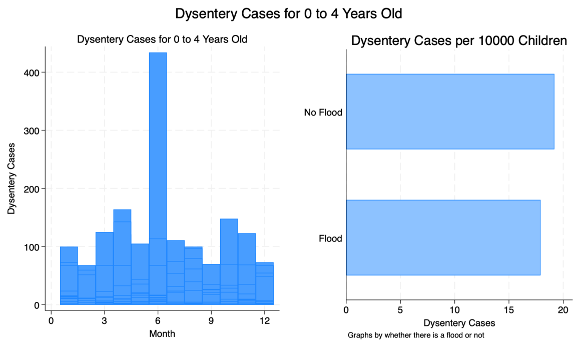


Figure 5: Dysentery cases

Table 1: Descriptive Statistics for Variables by Flood Status

| | (1) Flood mean | (2) No Flood mean | difference | (3) Difference t-statistic | p-value |
|---|---|---|---|---|---|
| Malaria confirmed | 1966.4215 | 1591.5849 | -374.84 | -9.15 | 0.00 |
| Cholera (HIA1) | 0.1235 | 0.0473 | -0.08 | -3.02 | 0.00 |
| Typhoid | 0.1440 | 0.2530 | 0.11 | 1.98 | 0.05 |
| Dysentery | 17.9163 | 19.1886 | 1.27 | 1.51 | 0.13 |
| Total rainfall (in mm) | 199.3647 | 78.1642 | -121.20 | -210.19 | 0.00 |
| Temperature (in $^{\circ}C$) | 28.0033 | 28.3870 | 0.38 | 24.87 | 0.00 |
| Mean elevation (in metres) | 1119.5222 | 1143.1373 | 23.62 | 22.35 | 0.00 |
| Distance to lake (in Km) | 243.9180 | 215.8317 | -28.09 | -36.11 | 0.00 |
| Distance to river (in Km) | 213.9020 | 219.5920 | 5.69 | 7.62 | 0.00 |
| Nighttime lights | 4.1143 | 9.5165 | 5.40 | 57.46 | 0.00 |
| Total number of facilities | 2.4080 | 2.5262 | 0.12 | 11.37 | 0.00 |
| *N* | 39081 | 333476 | 372557 | | |

Note: Diseases cases are per 10000 of under 5 years old population. Cholera (HIA1) - "Health Impact Assessment" (HIA1) is a reporting tool used in health facilities.

## 6.2 Regression Results

I estimate equation (1) above using a fixed effects linear probability model with flood as a dummy variable, and the results are reported in Tables 2 to 5 below, with the fixed effects regressions in the last two columns. In column (1) of each of the tables, I show what the estimate on the flood variable looks like without controls. Here, we can see that there are some endogeneity issues at play, specifically in Tables 1 and 4, where it appears that floods reduce the incidence of cholera and typhoid, respectively. With the addition of controls, some of this endogeneity in the form of omitted variable bias is tackled, but it is not complete. Fixed effects regressions help with isolating a causal impact of floods on the incidence of the disease as they control for some ward-specific time invariant factors, and some factors that move with time that could affect floods and the incidence of diseases, such as if in one year, there is a cholera outbreak and certain measures are taken to limit the spread of the disease.

In columns (3) and (4), I present the results of the fixed effects regressions, which are what we are interested in. Column (3) is our preferred estimation strategy, with column (4) offering an alternative and much stricter (or conservative) estimation of the fixed effects model from equation (1). The results below in Tables 2 and 3 for cholera and malaria, respectively, show that the probability of recording cases of these diseases increases when it floods, and this is consistent with what is observed in existing literature on flooding and diseases as evidenced by (Boyce et al., 2016; Deshpande et al., 2020; Picarelli et al., 2017; Suriya & Mudgal, 2012). For cholera, this probability is higher by 0.00083 or 0.083 percentage points when it floods and is statistically significant, although this significance disappears when we switch fixed effects from ward and year fixed effects to constituency by year and ward fixed effects. Taking the sample average baseline probability of 0.00112, the increase in the probability that facilities in wards where it floods record a cholera case in a given month and year represents a 74 percent increase in probability. For malaria, this probability of recording a malaria case when it floods is higher by 0.040 or 4 percentage points, and statistically significant regardless of the fixed effects used. With a baseline sample average probability of 0.795, the increase in probability by 0.040 shows that flooding increases the likelihood of recording malaria cases by about 5 percent.

In Tables 4 and 5, we see that floods do not have an impact on the probability that facilities record typhoid or dysentery cases. These results may seem to deviate from results reported in existing literature where flooding increased the occurrence of dysentery in Xinxiang city in Henan province in China (Ni et al., 2014), or where heavy rains were found to increase the incidence of diarrhea diseases (Deshpande et al., 2020). It is worth noting, however, that the typhoid and dysentery cases in our data are "suspected" or "clinical" cases. There are no tests conducted to confirm dysentery and typhoid cases, so there is a possibility that some of these cases are sometimes categorized as "diarrhea". Indeed, this aspect of "suspected" cases is something that some district health officers cautioned on when using the dysentery and typhoid data - to take the cases as "suspected" rather than confirmed.

Table 2: Incidence of Cholera in Children

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood (= 1) | -0.00070*** | 0.00015 | 0.00083* | 0.00034 |
| | (0.00020) | (0.00054) | (0.00048) | (0.00047) |
| Observations | 372,557 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.000 | 0.003 | 0.025 | 0.054 |
| Other controls | No | Yes | | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.00139 | 0.00112 | 0.00112 | 0.00112 |

Robust standard errors are in parentheses and are clustered at the ward level

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Incidence of Malaria in Children

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood (= 1) | 0.04021*** | -0.04117*** | 0.04045*** | 0.05542*** |
| | (0.00586) | (0.01025) | (0.00695) | (0.00706) |
| Observations | 372,557 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.001 | 0.105 | 0.378 | 0.409 |
| Other controls | No | Yes | Yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.801 | 0.795 | 0.795 | 0.795 |

Robust standard errors are in parentheses and are clustered at the ward level

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Incidence of Typhoid in Children

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood (= 1) | -0.00321*** | -0.00013 | 0.00066 | 0.00038 |
| | (0.00070) | (0.00078) | (0.00076) | (0.00079) |
| Observations | 372,557 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.000 | 0.014 | 0.083 | 0.106 |
| Other controls | No | Yes | Yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.00536 | 0.00433 | 0.00433 | 0.00433 |

Robust standard errors are in parentheses and are clustered at the ward level

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Incidence of Dysentery in Children

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood (= 1) | 0.03185*** | 0.02992*** | -0.00457 | -0.00419 |
| | (0.00442) | (0.00903) | (0.00676) | (0.00673) |
| Observations | 372,557 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.001 | 0.032 | 0.257 | 0.293 |
| Other controls | No | Yes | yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.172 | 0.225 | 0.225 | 0.225 |

Robust standard errors are in parentheses and are clustered at the ward level

*** p<0.01, ** p<0.05, * p<0.1

Given we have the number of floods in a month for a given year in each ward, I also examine whether the incidence of diseases changes with the number of floods. Thus, I re-estimate equation (1) above with $Flood_{wt}$ as a continuous count variable. The results for this regression for the incidence of cholera, malaria, typhoid, and dysentery are presented in Tables 6 to 9 below. Here, we see that while the number of floods does not have an effect on the probability that a ward records a cholera, typhoid, or dysentery case, one more flood will increase the probability of the incidence of malaria by 0.001 or 0.1 percentage points, and this effect is statistically significant. These results on flooding impacts on the incidence of disease entail that with waterborne diseases, one more flood may not matter - so long as it floods, the incidence of cholera will increase. For malaria, however, how many floods a ward experiences will matter for the incidence of malaria in that ward for children under the age of five years.

Table 6: Incidence of Cholera in Children: Effect of Number of Floods

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood count | 0.00001 | 0.00003 | 0.00006 | 0.00004 |
| | (0.00002) | (0.00004) | (0.00004) | (0.00004) |
| Observations | 277,139 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.000 | 0.003 | 0.025 | 0.054 |
| Other controls | No | Yes | Yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.000805 | 0.00112 | 0.00112 | 0.00112 |

Robust standard errors are in parentheses and are clustered at the ward level
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Incidence of Malaria in Children: Effect of Number of Floods

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood count | 0.00315*** | -0.00255*** | 0.00130** | 0.00237*** |
| | (0.00066) | (0.00096) | (0.00058) | (0.00060) |
| Observations | 277,139 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.001 | 0.105 | 0.378 | 0.409 |
| Other controls | No | Yes | Yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.806 | 0.795 | 0.795 | 0.795 |

Robust standard errors are in parentheses and are clustered at the ward level
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Incidence of Typhoid in Children: Number of Floods

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood count | -0.00018*** | -0.00002 | 0.00006 | 0.00004 |
| | (0.00005) | (0.00005) | (0.00004) | (0.00005) |
| Observations | 277,139 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.000 | 0.014 | 0.083 | 0.106 |
| Other controls | No | Yes | Yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.00444 | 0.00433 | 0.00433 | 0.00433 |

Robust standard errors are in parentheses and are clustered at the ward level

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Incidence of Dysentery in Children: Number of Floods

| VARIABLES | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Flood count | 0.00196*** | 0.00352*** | 0.00053 | 0.00081 |
| | (0.00063) | (0.00091) | (0.00054) | (0.00053) |
| Observations | 277,139 | 188,439 | 188,438 | 188,438 |
| R-squared | 0.000 | 0.032 | 0.257 | 0.293 |
| Other controls | No | Yes | Yes | Yes |
| Year FE | No | No | Yes | No |
| Ward FE | No | No | Yes | Yes |
| Constituency-year FE | No | No | No | Yes |
| Mean of the dependent variable | 0.188 | 0.225 | 0.225 | 0.225 |

Robust standard errors are in parentheses and are clustered at the ward level

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The incidence of diseases is influenced by many factors, and one of those is the development of an area. Therefore, it is important to analyze whether the results we are seeing are being driven by the urban population or if the effect of flooding affects both the rural and urban populations equally. While the data does not have designations of rural and urban areas, I mimic this analysis by looking at how flooding impacts the incidence of diseases, whether a facility in a ward is located in a main town or not. Main towns are provincial capitals and tend to be more developed than non-main towns. In this analysis, the main towns are: Lusaka for Lusaka province, Ndola for Copperbelt province, Solwezi for North-Western province, Choma for Southern province, Mansa for Luapula province, Kasama for Northern province, Chinsali for Muchinga province, Mongu for Western province, Kabwe for Central province, and Chipata for Eastern province. I also added Kitwe and Livingstone, as they are considered cities and should have the same level of development as the main towns. Results of the regressions by interacting town status with the flood variable are reported in Table 10 below. I only report results for diseases where we saw an impact of flooding on the incidence, as shown in Tables 2 to 4, i.e., on cholera and confirmed malaria cases.

Table 10: Incidence of Diseases in Children (By Main Town)

| VARIABLES | Cholera (HIA1) | Cholera (HIA1) | Malaria | Malaria |
|---|---|---|---|---|
| Flood (= 1) | 0.00085* | 0.00038 | 0.04119*** | 0.05656*** |
| | (0.00046) | (0.00046) | (0.00706) | (0.00715) |
| Flood ( = 1) * main town | -0.00017 | -0.00047 | -0.00804 | -0.01249 |
| | (0.00064) | (0.00061) | (0.00812) | (0.00780) |
| Observations | 188,438 | 188,438 | 188,438 | 188,438 |
| R-squared | 0.02458 | 0.05356 | 0.37838 | 0.40944 |
| Other controls | Yes | Yes | Yes | Yes |
| Year FE | Yes | No | Yes | No |
| Ward FE | Yes | Yes | Yes | Yes |
| Constituency-year FE | No | Yes | No | Yes |
| Mean of the dependent variable | 0.00112 | 0.00112 | 0.795 | 0.795 |

Robust standard errors are in parentheses and are clustered at the ward level
*** p<0.01, ** p<0.05, * p<0.1

From the results in Table 10, the health impacts of flooding do not differ between main towns and non-main towns. This shows that the child health impacts of flooding are shared across all health facilities in flooded and flood-prone wards rather than where economic activity and development may be concentrated.

# 7 Conclusion

Using satellite flood data from the MODIS Water Product and news articles confirming the occurrence of floods in certain months, I examine how floods impact the incidence of cases of cholera, malaria, typhoid, and dysentery in children under the age of five years in Zambia from

2012 to 2020. I find that for cholera and malaria, floods increase the incidence of these diseases as facilities in wards have a higher probability of recording cases of cholera and malaria when it floods than when it does not flood, and the results are statistically significant (although for statistical significance for incidence of cholera disappears with the addition of constituency by year fixed effects). For typhoid and dysentery, I do not find any effects of flooding on the incidence of the diseases. The results of the regressions in this analysis on malaria and cholera are consistent with the literature, while dysentery and typhoid give a different perspective on how flooding impacts diseases classified as "diarrhea" diseases.

My analysis does not lump waterborne diseases under "diarrhea" diseases but estimates each of the health outcomes separately. Thus, differences in results with existing literature can occur, but the results still offer insight into how year-to-year floods, small and big, affect the prevalence of different types of diseases known to affect children, i.e., malaria, and waterborne diseases such as cholera, typhoid, and dysentery. Thus, sensitization efforts on the detrimental effects of flooding on malaria and cholera can help mitigate the increased incidence of malaria and cholera. Targeting all "diarrhea" or "waterborne" diseases against the health effects of flooding for children under the age of five years old risks lumping diseases that are not necessarily affected by flooding.

The results from this research also show that the child health impacts of flooding are not restricted to main town areas but are experienced all across the country, demonstrating that the effects of flooding are both an urban and rural problem. The results presented in this paper are a first step to quantifying the effects of flooding on child health outcomes in Zambia, and possible next steps will include more heterogeneity analysis on age and location (other than town versus non-town), and why floods do not impact the incidence of typhoid and dysentery.

# References

Berazneva, J., & Byker, T. S. (2017). Does forest loss increase human disease? evidence from nigeria. *American Economic Review*, *107*(5), 516–521. https://doi.org/10.1257/aer.p20171132

Berazneva, J., & Byker, T. S. (2024). Impacts of environmental degradation: Forest loss, malaria, and child outcomes in nigeria. *The Review of Economics and Statistics*, *106*(5), 1254–1267. https://doi.org/10.1162/rest_a_01238

Berggreen, S., & Mattisson, L. (2024). The curse of bad geography: Stagnant water, diseases, and children's human capital.

Boyce, R., Reyes, R., Matte, M., Ntaro, M., Mulogo, E., Metlay, J. P., Band, L., & Siedner, M. J. (2016). Severe flooding and malaria transmission in the western ugandan highlands: Implications for disease control in an era of global climate change. *The Journal of Infectious Diseases*, *214*(9), 1403–1410. https://doi.org/10.1093/infdis/jiw363

Chandir, S., Glennerster, R., Haroon, M., Jee, E., & Siddiqi, D. A. (2023). Does childhood immunization rebound after extreme shocks? evidence from floods and strikes in pakistan. *AEA Papers and Proceedings*, *113*, 642–646. https://doi.org/10.1257/pandp.20231110

Chen, X., & Nordhaus, W. D. (2010, August). The value of luminosity data as a proxy for economic statistics. https://doi.org/10.3386/w16317

Deshpande, A., Chang, H. H., & Levy, K. (2020). Heavy rainfall events and diarrheal diseases: The role of urban–rural geography. *The American Journal of Tropical Medicine and Hygiene*, *103*(3), 1043–1049. https://doi.org/10.4269/ajtmh.19-0768

Elsanousi, Y. E. A., Elmahi, A. S., Pereira, I., & Debacker, M. (2018). Impact of the 2013 floods on the incidence of malaria in almanagil locality, gezira state, sudan. *PLoS currents*, *10*, ecurrents.dis.8267b8917b47bc12ff3a712fe4589fe1. https://doi.org/10.1371/currents.dis.8267b8917b47bc12ff3a712fe4589fe1

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas [_eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5086]. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., & Verdin, A. P. (2014). *A quasi-global precipitation time series for drought monitoring* (ISSN: 2327-638X). US Geological Survey. https://doi.org/10.3133/ds832

Garg, T. (2019). Ecosystems and human health: The local benefits of forest cover in indonesia. *Journal of Environmental Economics and Management*, *98*, 102271. https://doi.org/10.1016/j.jeem.2019.102271

Levy, K., Woster, A. P., Goldstein, R. S., & Carlton, E. J. (2016). Untangling the impacts of climate change on waterborne diseases: A systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought [Publisher: American Chemical Society]. *Environmental Science & Technology*, *50*(10), 4905–4922. https://doi.org/10.1021/acs.est.5b06186

Li, X., Zhou, Y., zhao, M., & Zhao, X. (2021). Harmonization of DMSP and VIIRS nighttime light data from 1992-2020 at the global scale [Artwork Size: 1114861857 Bytes Pages: 1114861857 Bytes]. https://doi.org/10.6084/M9.FIGSHARE.9828827.V5

Musamba, C. D., & Phiri, K. (2019). Understanding the usage of constituency development fund (CDF) in zambia: The case of education, health, water and sanitation projects [Publisher: Jesuit Centre for Theological Reflection]. Retrieved December 16, 2024, from https://repository.jctr.org.zm/handle/20.500.14274/237

Mwaba, J., Debes, A. K., Shea, P., Mukonka, V., Chewe, O., Chisenga, C., Simuyandi, M., Kwenda, G., Sack, D., Chilengi, R., & Ali, M. (2020). Identification of cholera hotspots in zambia: A spatiotemporal analysis of cholera data from 2008 to 2017. *PLoS Ne-*

*glected Tropical Diseases*, *14*(4), e0008227. https://doi.org/10.1371/journal.pntd.0008227

Ni, W., Ding, G., Li, Y., Li, H., Liu, Q., & Jiang, B. (2014). Effects of the floods on dysentery in north central region of henan province, china from 2004 to 2009 [Publisher: Elsevier]. *Journal of Infection*, *69*(5), 430–439. https://doi.org/10.1016/j.jinf.2014.05.016

Okaka, F. O., & Odhiambo, B. D. O. (2018). Relationship between flooding and out break of infectious diseasesin kenya: A review of the literature [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10 *Journal of Environmental and Public Health*, *2018*(1), 5452938. https://doi.org/10.1155/2018/5452938

Pagel, J. (2025). A natural resource curse: The unintended effects of gold mining on malaria. *Ecological Economics*, *230*, 108469. https://doi.org/10.1016/j.ecolecon.2024.108469

Patel, D. (2023, November 18). Floods. Retrieved June 15, 2025, from https://papers.ssrn.com/abstract=4636828

Picarelli, N., Jaupart, P., & Chen, Y. (2017). Cholera in times of floods1 weather shocks & health in dar es salaam, reference number: C-40404-tza-1.

Sasaki, S., Suzuki, H., Fujino, Y., Kimura, Y., & Cheelo, M. (2009). Impact of drainage networks on cholera outbreaks in lusaka, zambia. *American Journal of Public Health*, *99*(11), 1982–1987. https://doi.org/10.2105/AJPH.2008.151076

Suhr, F., & Steinert, J. I. (2022). Epidemiology of floods in sub-saharan africa: A systematic review of health outcomes. *BMC Public Health*, *22*(1), 268. https://doi.org/10.1186/s12889-022-12584-4

Suriya, S., & Mudgal, B. V. (2012). Impact of urbanization on flooding: The thirusoolam sub watershed – a case study. *Journal of Hydrology*, *412-413*, 210–219. https://doi.org/10.1016/j.jhydrol.2011.05.008

theigc.org